DOCUMENT RESUME

ED 036 177 24 EM 007 763

AUTHOR DENO, STANLEY N.; JENKINS, JOSEPH R.
TITLE EVALUATING PRE-PLANNED CURRICULUM OBJECTIVES.
INSTITUTION RESEARCH FOR BETTER SCHOOLS, INC., PHILADELPHIA, PA.
SPONS AGENCY OFFICE OF EDUCATION (DHEW), WASHINGTON, D.C. BUREAU OF RESEARCH.
BUREAU NO BR-6-2867
PUB DATE 67
CONTRACT OEC-1-7-062867-3053
NOTE 45P.

EDRS PRICE EDRS PRICE MF-$0.25 HC-$2.35
DESCRIPTORS *BEHAVIORAL OBJECTIVES, EVALUATION CRITERIA, EVALUATION METHODS, *INDIVIDUALIZED INSTRUCTION, *MEASUREMENT TECHNIQUES
IDENTIFIERS INDIVIDUALLY PRESCRIBED INSTRUCTION, IPI

ABSTRACT

A BASIC REQUIREMENT IN CURRICULUM DESIGN AND DEVELOPMENT IS THAT CURRICULUM OBJECTIVES BE STATED IN TERMS OF OBSERVABLE HUMAN BEHAVIOR. THIS PROJECT DEVELOPED A TECHNIQUE FOR EVALUATING CURRICULUM OBJECTIVES BASED UPON THE ANALYSIS OF BEHAVIORAL OBJECTIVES AS CLASSIFIED INTO THREE COMPONENTS--ACTION WORD, CONTEXT, AND CRITERION. IN ORDER TO DETERMINE WHETHER A PERSON TRAINED TO EVALUATE CURRICULUM OBJECTIVES IN TERMS OF SPECIFIED CRITERIA COULD DO SO CONSISTENTLY, A GROUP OF PUBLIC SCHOOL TEACHERS USED THE CRITERION INSTRUMENT TO ANALYZE A STATED OBJECTIVE. THE RESULTS OF A STUDY OF THEIR EFFORTS ARE REPORTED. THE EXTENT TO WHICH THE OBJECTIVES OF THE MATHEMATICS CONTINUUM USED IN INDIVIDUALLY PRESCRIBED INSTRUCTION (IPI) CONFORM TO THE STATED GOAL IS ALSO DISCUSSED. THE MATERIAL USED TO COLLECT INFORMATION IS APPENDED, ALONG WITH SOME OF THE RAW DATA. (JY)

EVALUATING PRE-PLANNED CURRICULUM OBJECTIVES

Research for Better Schools, Inc.

Stanley R. Deno, Principal Investigator
Joseph R. Jenkins, Co-Principal Investigator

University of Delaware
Newark, Delaware

EVALUATING PRE-PLANNED CURRICULUM OBJECTIVES

Research for Better Schools, Inc.

Stanley L. Deno, Principal Investigator
Joseph R. Jenkins, Co-Principal Investigator

University of Delaware
Newark, Delaware

1967

The research reported herein was supported by
the Research for Better Schools, Inc., Philadelphia, Pa.

EVALUATING PRE-PLANNED CURRICULUM OBJECTIVES

INTRODUCTION

If there is a single most important influence which behavioristic psychology has had upon curriculum design and development, it is perhaps the requirement that curriculum objectives be stated in terms of observable human behavior. The position taken by educational psychologists with a behavioristic or neo-behavioristic orientation is that one can determine the conditions most appropriate for learning only after the terminal behavior has been specified. Once terminal objectives are specified in performance terms, it is then possible to proceed to develop instructional sequences appropriate for promoting the acquisition of that behavior.

Curriculum developers who accept the premise that objectives should be stated in terms of observable human performance argue that only in this way may achievement of objectives be determined and the effectiveness of a particular curriculum sequence be assessed (Taber, Glaser and Schaefer, 1965; Gagné, 1965).

Purpose:

While many have argued the merits of specifying curriculum objectives in behavioral terms, the present study was not undertaken to support the contention that curriculum objectives be stated behaviorally. Rather, the present project was conducted for the following reasons:

1. To develop a set of criteria for determining the extent to which curriculum objectives had been stated in terms of observable student behavior, and to specify a rationale for the criteria and technique.
2. To train a group of public school teachers to use the criterion instrument to analyze a stated objective of a pre-planned curriculum.
3. To determine whether a person trained to evaluate curriculum objectives in terms of specified criteria could do so consistently.

4. To determine the extent to which the objectives of the mathematics continuum used in Individually Prescribed Instruction are stated in behavioral terms.

Background:

The fundamental principle guiding the activities of the project was that whatever technique would be adopted to determine the behaviorality of stated curriculum objectives, it should be simple enough to be reliably and effectively used by people who were neither curriculum nor subject matter specialists.

A number of pilot studies were conducted to determine what approach would be most effective. These preliminary efforts revealed that typical pre-service and in-service teachers tended not to think of educational outcomes in terms of human behavior. They tended, rather, to equate dispositional states with their behavioral concomitants. When presented with objectives varying in degree of behavioral specificity, and asked the question, "Which of these objectives most clearly describes observable human behavior?", pilot subjects responded very nearly randomly or arbitrarily, and complained they did not understand. Even after attempts to clarify the question, subjects failed to reliably respond. These preliminary efforts confirmed our suspicions that any workable technique for determining the behaviorality of objectives would probably involve an analysis of behavioral objectives into their critical components.

The prior analyses of instructional objectives by Mager (1962), Miller (1961), Gagné (1965) and Taber, Glaser and Schaefer (1965), served as a prototype for the present analysis. A review of these writings indicated that while there was some variability in the recommended number of components in a behavioral objective, there were three components upon which all agreed. These were:

1. A specification of the kind of behavior which the learner is expected to perform; that is, an action word or verb.
2. A description of the important conditions under which the behavior will be expected to occur; that is, a context (or signal) in which the action takes place.
3. A description of the criterion for the kind of behavior which is to be performed; that is, a statement of how well the learner must perform to have his behavior accepted.

Some users of behavioral objectives have added a fourth component to the list, and while at least one well-known educational psychologist (Gagné, 1965) has attempted to find commonality among these such a synthesis is not clearly feasible. Further, bringing in a fourth component, such as an object acted upon (Gagné), complicates an analysis since some degree of expertise in the language of a particular subject matter is required. This, of course, would have violated our guiding principle that the evaluation technique be useable by personnel relatively unsophisticated in the subject matter.

Locus of Behaviorality. Inspection of the three components led to the immediate conclusion that behaviorality resides in the first component: "The kind of behavior which the learner is expected to perform." Attention must be focused on the character of the verbs contained in behavioral objectives in order to estimate behaviorality since it is the verb or action word of an objective which describes behavior.

Curriculum developers advocating the use of behavioral objectives have been adamant that objectives contain a description of specific observable human behavior. However, an examination of the words recommended by these curriculum developers reveals that what is meant by behaviorality is "observability" (for example, "to write", "to recite", "to list", "to state", rather than "to know", "to understand", "to appreciate", Mager, 1962). A

close look at many of the recommended words shows that all terms are, in some sense, equally identifiable as descriptions of behavior; however, not all can be said to denote directly observable behavior (for example, "to identify", "to differentiate", "to solve", "to compare", Mager, 1962; or, "to distinguish", "to demonstrate", Commission on Science Education, 1963). When the behavioristic curriculum developer demands an answer to the question,"What will the learner be doing when he is demonstrating that he has achieved the objective?", he seems to be saying to us, "What observable data are necessary to make the inference that learning has occurred?".

Another way of describing the problem of describing behavior consistently has been proposed by Mandler and Kessen (1959). In a careful analysis of the language of science those authors state that the criterion for a term's acceptability within science is its invariant usage; that is, its consistent usage by all members of a particular language community. If a term is used by all members of the same linguistic community invariantly, then, apparently, it is safe to assume that for that linguistic community the referent is clear. In other words, the presence or absense of a particular event (referent) may be inferred from the consistency with which members of a particular language community use a term. Thus, when ten physicists use the same term "proton" at the same time, the inference to be made is that there is a consistent referent for that term.

It occurred to the present investigators that whatever is present in the external world of the curriculum developer is present regardless of what terms are used in their attempts to describe that world. What is important is that curriculum developers use the same terms at the same time to apply to the same events so that agreement on outcomes of the instruction can be achieved.

Apparently, the concern of those promoting objectives which include specific observable human behavior is that terms such as "know", "understand"

and "appreciate" are not used invariantly by native speakers of the English language, while words such as "write", "identify" and "solve" are used invariantly. The presumed variance in usage of a term such as "know" would mean by implication that the stimulus occasion for using such a term is not invariant--that there is no event consistently labelable with the term "know". By equating behaviorality with observability curriculum developers have, apparently, attempted to achieve a high degree of invariance with respect to usage. In this way, equivocation over the successful attainment of objectives is to be eliminated.

Prior to our analysis of the action word or verb in terms of invariance, it had seemed reasonable that certain behaviors could be categorized as occurring or not occurring simply by observing them (e.g., "writing" or "saying"). Other behaviors thought by some to be directly observable, such as "to identify" or "to classify", apparently were actually processes which were essentially as unobservable as "know" and "appreciate". Subsequent consideration led to the conclusion, however, that even actions such as "write" and "say" would sometimes not be identified invariantly, and that pursuit of the line of reasoning that one could tabulate the occurrence or non-occurrence of a behavior simply by watching the organism would lead to the same difficulties encountered by behavioristic psychologists attempting to define the bar pressing response. Skinner avoids this complication by attending not to the topography of the response but to the effect of the response upon the environment. Thus, the bar pressing response is defined in terms of an electrical contact made, regardless of the manner (or form of behavior) by which the depression of the bar occurred (e.g., pushing with snout, tail or forepaws).

Following Skinner's lead, our next attempt was to develop a technique whereby our teacher-judges would categorize the action word of an objective in terms of its effect upon the environment. As with the earlier attempt to

rate the behaviorality of an entire objective this effort was remarkably unsuccessful. A question such as, "To what extent would you be able to observe the effect of the word (action word)?" produced seemingly random behavior and misunderstanding. Apparently it would be necessary to describe the effect of the behavior as well as the behavior itself in order for subjects to make a judgement of this type.

In a behavioral objective the effect of a particular response is contained in the criterion, and our primary task here was to determine the behaviorality of the action words used. Since this approach tended to take us away from the behavior and focused our attention on the criterion instead, it too was abandoned.

Rationale:

The pilot work led us to develop a technique for evaluating curriculum objectives based upon the analysis of behavioral objectives into the three components upon which all writers agreed - action word, context and criterion.

Action Word. Since it appeared to us that most task analysts and curriculum developers were equating behaviorality with observability in an effort to achieve invariance, our solution to the problem of determining behaviorality was to structure the task for our subjects as one of observation. The question, "To what extent could you directly observe the behavior described by the term (action word)?" was a question which seemed to produce meaningful and consistent responding by all pre-service and in-service teachers. Our assumption was that judgements of "direct observability" were an indication not only of the extent to which a term served to label observable behavior, but also (and perhaps more importantly) the degree to which a term was invariantly used by a native speaker.

Since we were primarily concerned with analyzing the behaviorality of instructional objectives of the type recommended by curriculum developers with a behavioral bent it seemed appropriate to construct a scale within the limits

established by the experts in this field. Specifically, the polar extremes of observability have, essentially, been established in the writings of Miller, Gagné, and Mager. Maximum observability is implied by words such as "to underline" or "to point to", while maximum unobservability is implied by words such as "to know", "to understand" and "to appreciate".

Our prediction was as follows:

If the polar extremes of a continuum of observability are established for a group of native speakers of the English language, and those speakers are asked to locate a set of labels (action words) on that continuum, then, words such as "write", "say" and "underline" will tend to fall toward the "easily observed" pole, while words such as "know", "understand" and "appreciate" will tend to fall toward the "difficultly observed" pole. Further, our prediction was that many of the words commonly used and recommended by curriculum developers in writing behavioral objectives (such as "identify", "solve" and "distinguish") would fall at an intermediate position on the continuum.

The class of words falling at intermediate positions would be particularly interesting since they could, in no sense, be thought of as directly observable, would probably be used less invariantly and finally would be subject to greater disagreements than terms rated as observable. With such a scale an attempt was made to determine the behaviorality of instructional objectives.

Context or Signal. Following specification of the behavior that the learner is expected to perform as part of an objective, the conditions under which that behavior is expected to occur must be described (Taber, Glaser and Schaefer). Or, as others have suggested, an indicator or signal which is to be the discriminative stimulus for the behavior (Miller) or, "words denoting the stimulus situation which initiates the performance" (Gagné) must be specified.

As with the action verb component of a behavioral objective there is, or can be made, a differentiation among different levels of generality of the

contextual or stimulus component. For example, the context "when given the following set of addition problems: 2 + 2 = _____, 2 + 3 = _____, 3 + 5 = _____, and the direction 'write the sums of the above set of addition problems in the blank spaces'" is a more specific context than "when given single digit addition problems". Clearly, the more general context of the latter implies that any of a class of problems may by given to the student for him to act upon (although it is not entirely clear whether those problems will be in the form of mathematical sentences or in columns with a plus sign beside each number and a line below).

While the first more specific type of context is open to less interpretation it does not suggest what inference should be drawn about the pupil's capability with respect to a general class of problems. Characteristic of the more specific context is the inclusion of the actual test question or test items as a part of the context. . The more general context would serve as a basis for generating a wide variety of problems which, presumably, the learner ought to be able to perform.

The distinction between general and specific context is made here because it seemed to the present investigators that when curriculum developers writing behavioral objectives demanded "specific observable human behavior," oftentimes the specificity which is sought is determined not only by the action verb but also by the nature of the context involved. If the context is specific then the objective will appear to be more specific (and, perhaps to some, more behavioral or observable since the objective has been described in greater detail.

Criterion. The final component of a behavioral objective upon which there is general agreement is that of the criterion of acceptable performance. An examination of both the literature dealing with instructional objectives and available curriculum objectives revealed that there are generally two types of criteria employed by most curriculum writers. The first type of criterion

is that which is established by explicitly stating the correct answer to a given objective or problem. For example, in the objective "when given the problem: 5 + 2 = _____, the student writes 7 in the blank", the criterion for an acceptable response is explicitly given as "7". Anyone attempting to score pupil achievement on the objective would be able to do so because the problem and its acceptable answer "7" is explicitly stated.

The second type of criterion frequently occurring is that which indicates some percentage or proportion of correct responding on the part of the student. This second type of criterion is exemplified in the objective "counts orally from one to ten at least six out of seven times correctly." That the student counts from one to ten once correctly may be unsatisfactory for some. To have his response accepted the student is required to perform correctly six or seven times, and at that time the criterion of acceptability can be applied. Another example of this type might be, "when given a multiple choice examination on the nature of plastics the student will answer 90% of the questions correctly." Again a proportionality of total responding is specified as an accepted level of performance.

It is easy to ignore the often made recommendation that a criterion for acceptability be explicitly stated by saying that a criterion is nearly always implied in a stated objective. Interestingly enough, however, the criterion that is obviously implied often is not the criterion that is emplpyed in an instructional situation.*

---

* Although the objectives from the IPI math continuum seem to imply a criterion of 100% correct responses, users of this curriculum employ a criterion of 80% correct (Dover Public Schools).

Conclusion:

The purpose of the present investigation was to determine whether a technique for analyzing behavioral objectives in terms of the critical components outlined above could be developed and reliably used by a typical group of in-service teachers.

We believe that it is most important to note at this time that the above description of the components of behavioral objectives and the observation of behaviorality and specificity are made neither prescriptively nor evaluatively. Rather, an analysis of this type was undertaken to develop a technique for determining the "behaviorality" of curriculum objectives which could be both reliably and meaningfully used. In a sense, then, it is more appropriate to refer to this rationale and the subsequent procedures involved as an analysis rather than an evaluation.

It might be reasonable to conclude, however, that should one wish to reliably communicate the objectives of a curriculum, the more specific the content, observable the behavior, and carefully stated the criterion, the more likely are people to agree that a desired objective has been achieved. On the other hand, as an objective becomes more specific and "behavioral" it becomes less inclusive. This increased specificity, then, produces objectives which may seem superficial or educationally trivial. Thus, increased reliability of communication may be purchased at the risk of sacrificing linguistic sophistication and educational respectability.

PROCEDURES

Consistent with the decision to analyze the objective in terms of critical components, as described in the rationale, a set of instructional objectives were classified in terms of context, specificity, and criterion. However, the problem encountered in attempting to rate the behaviorality of an action word which is embedded within a completely stated objective (the "observableness" of the behavior named by the verb, is often confused with the specificity of the context), resulted in the decision to perform the analysis in two parts. First, the action words from IPI along with a larger set of words commonly found in educational objectives were rated on an observability dimension. Second, a sample of objectives from the IPI math continuum were classified in terms of the signal and criterion components.

Subjects

Thirteen in-service teachers and one principal participated at one time or another in the activity involved in the analysis of the instructional objective. Only 11 of those teachers and the principal were available during the entirety of the analysis. The subjects were teachers from both elementary and secondary schools ranging in grade level from first to twelfth.

Determination of Observability

The extent to which a verb labels an observable event was determined by having subjects judge or rate the event on a scale of observability. This scale was constructed with five points, ranging from presumably "easily observable" events such as to hit and to bite, to "difficultly observable" events such as to believe and to sympathize.

Since the purpose of the present investigation was to determine the behaviorality of the IPI objectives nearly all of the verbs contained in

working paper #20 of the Individually Prescribed Instruction Mathematics Curriculum (April 1, 1967) were extracted and included for rating. In all, 45 different action words were taken from the IPI mathematic objective. As previously mentioned, a number of additional verbs often used in instructional objectives were included in a set of verbs to be rated. These verbs (54) brought the total number of verbs rated to 99.

The ninety-nine action words were typed in a single double-spaced column on four pages. In addition, a set of written instructions describing the nature of the task for the rater, a general description of the problem, and examples of the polar extremes of the rating scale served as a cover sheet to the four pages of verbs. The instructions along with the four pages of verbs served as a test booklet which was then distributed to each subject. The page order was randomized to reduce the effect of particular sequences upon the ratings given by the subject. A sample test booklet is included as Appendix A of this report. In addition to the written instructions on the cover sheet of the booklet the students were requested not to return to a page after it was once completed.

The subjects performed their ratings at the same time in a group. Fifteen minutes wereallowed for completion of the rating. All subjects easily completed the rating within the fifteen minute time period.

Component Classification

Classification of objectives on the basis of critical components was accomplished by providing the students with a prepared description of the components. This description of the components is included as Appendix B to this report. The description of critical components was to serve as the only instruction given to the subject as to how the objectives were to be classified. After reading the description the students were required to

classify a set of objectives randomly selected from individually prescribed instruction mathematics continuum. Subjects were provided with a classification sheet in which to indicate their choice of category.

Prior to the classification task it became apparent that many of the objectives from the IPI continuum were compound or complex in nature. That is, very often more than one operation or action is called for from the student, and more than one context is included. In effect, what appears to be a single objective often turns out to be two or more objectives. Consequently, subjects involved in the analysis were directed to classify each objective only in terms of the first operation described.

Subsequent to reading the instructions the subjects were given a randomly selected set of IPI objectives (approximately 45) and a classification checklist. The subjects were requested to classify in terms of presence or absence of the critical components of those objectives. However, subjects expressed considerable consternation with their task after having rated approximately 20 of the objectives. The subjects stated that they were able to classify consistently at first, but as they proceeded they became confused. Consequently, rather than continue the subjects were requested to take home their descriptions for further study. In addition they were directed to generate one example of a general objective with context and criterion stated, and one example of the specific example with context and criterion stated.

On the following day the objectives that the students had constructed were printed on a chalk board and a practice session ensued where subjects attempted to determine the presence or absence of critical components in one another's objectives. Approximately two hours were spent during this practice session.

One day later a different set of 41 IPI objectives was randomly selected and presented to the subjects for critical component analysis. An unlimited amount of time was given for completion of the classification task, and all subjects completed the task within 45 minutes.

The results of both critical component classification tests are included in the following section.

RESULTS

The results are reported in terms of the two-part analysis regarding the observability of the verb in the component classification.

Verb Observability

The mean ratings and the variance of the ratings for the entire set of action words is contained in Table 1. The words are ranked from most observable to least observable. Appendix D contains the same action words ordered alphabetically with their respective mean ratings and variances.

TABLE 1

RANK-ORDER DISTRIBUTION OF MEANS FOR 99 VERB RATINGS

TERMS	MEANS	VARIANCES
to cover with a card	1.00000	0.00000
to lever press	1.00000	0.00000
to line-draw	1.00000	0.00000
to mark	1.08333	0.08333
to point to	1.08333	0.08333
to crossout	1.16666	0.33333
to underline	1.16666	0.33333
to walk	1.16666	0.33333
*to circle	1.25000	0.38636
to repeat orally	1.25000	0.38636
*to count orally	1.25000	0.75000
*to say	1.25000	0.75000
*to write	1.33333	0.42424
*to put on	1.41666	0.44696
*to read orally	1.50000	0.45454

TABLE 1 (continued)

TERMS	MEANS	VARIANCES
*to shade	1.50000	0.81818
to number	1.58333	0.44696
*to name	1.58333	0.81060
*to fill in	1.66666	0.96969
to label	1.75000	1.11363
*to state	1.75000	1.47727
*to remove	1.91666	0.62878
*to place	1.91666	0.99242
to tell what	1.91666	1.17424
*to draw	2.08333	0.99242
*to identify in writing	2.16666	1.42424
*to check	2.25000	1.29545
*to construct	2.25000	1.29545
*to match	2.33333	0.78787
*to take away	2.33333	1.15151
*to make	2.41666	0.99242
*to arrange	2.50333	0.62878
to finish	2.58333	0.62878
*to read	2.58333	0.81060
*to play	2.58333	1.71969
*to locate	2.66666	0.60606
*to connect	2.66666	1.15151
*to give	2.66666	1.33333
*to reject	2.75000	1.11363
*to select	2.75000	1.47727
to choose	2.83333	0.51515
*to partition	2.91666	0.44696
*to change	2.91666	0.99242
*to use	2.91666	1.17424
*to subtract	2.91666	1.35606
*to perform	3.00000	1.81818
*to total	3.00000	1.81818
*to divide	3.08333	0.81060
*to order	3.08333	0.99242
*to measure	3.08333	1.17424
*to add	3.08333	1.35606
*to supply	3.08333	1.35606
*to demonstrate	3.16666	0.87878
*to regroup	3.16666	1.06060
*to multiply	3.16666	1.24242
*to round off	3.16666	1.42424
*to group	3.25000	0.56818
*to complete	3.25000	0.93181
*to respond to	3.33333	0.60606
*to average	3.33333	1.17424
to summarize	3.33333	1.17424
to inquire	3.50000	0.81818
to utilize	3.50000	1.00000
*to borrow	3.58333	0.44696
to acknowledge	3.58333	1.17424

TABLE 1 (continued)

TERMS	MEANS	VARIANCES
*to find	3.66666	1.69696
*to identify	3.83333	0.37878
to see	3.83333	2.33333
*to convert	3.91666	1.35606
to distinguish	4.16666	0.87878
*to solve	4.25000	0.93931
*to apply	4.25000	1.11363
to develop	4.33333	0.42424
*to test	4.33333	0.42424
*to determine	4.33333	0.60606
to generate	4.33333	0.78787
*to create	4.33333	1.15151
to discriminate	4.50000	0.63636
*to recognize	4.58333	0.44696
to discover	4.75000	0.20454
to become competent	4.75000	0.38636
to infer	4.75000	0.38636
to like	4.75000	0.38636
to analyze	4.83333	0.15151
to be curious	4.83333	0.15151
to conclude	4.83333	0.15151
*to deduce	4.83333	0.15151
to feel	4.83333	0.15151
to concentrate	4.83333	0.33333
to perceive	4.83333	0.33333
to think	4.83333	0.33333
to think critically	4.83333	0.33333
to learn	4.83333	0.38636
to appreciate	4.91666	0.08333
to be aware	4.91666	0.08333
to know	4.91666	0.08333
to wonder	4.91666	0.08333
to realize fully	5.00000	0.00000
to understand	5.00000	0.00000

*Denotes IPI verbs.

Inspection of Table 1 shows that the verbs or action words extracted from IPI range in rate of observability from 1.3 (most observable) to 4.8 (least observable) with a median rating of approximately 2.8.

Since the variances are an indication of the consistency with which the various terms were rated, it is interesting to note that only 4 out of the 99 variances exceed 1.50.

Component Classification

The results of the first attempt at component classification are shown in Table 2. The left hand column denotes the level in the IPI continuum and the number of the IPI objectives under the level. The number in any box denotes the number of people classifying that component of the objective in the category indicated by the column heading. The surface consistency of the data obviates the necessity for applying any statistical test.

Table 2

Frequency Distribution of Subject's Ratings on Selected Objectives (first test).

Objective Number	Specific	General	Signal		Criterion	
			Stated	Not Stated	Stated	Not Stated
B-1-Numeration	2	8	10			10
B-2		10	10			10
B-3		10		10		10
B-4		10		10		10
B-5	1	9		10	1	9
B-6		10	1	9		10
B-7		10	1	9		10
E-1-Subtraction		10		10		10
E-2		10		10		10
E-3		10		10		10
E-1-Multiplication		10		10		10
E-2		10		10		10
E-3		10		10		10
E-4		10		10		10
D-1-Measurement		9		9		9
D-2		9		9	1	8
D-3		9		9		9
D-4		9		9	1	8
D-5		9		9	1	8
D-1-Geometry	1	8		9		9
D-2		9		9		9
D-5	2	7	1	8		9
D-6	1	7		8		8

The results of the component classification after the subjects have practiced is shown in Table 3. A total of 41 objectives from the IPI math continuum

were presented in the second test. Again there was virtually unanimous agreement on every classification of critical components for every one of the objectives.

Table 3

Frequency Distribution of Subject's Ratings on Selected Objectives (second test).

Objective Number	Specific	General	Signal		Criterion	
			Stated	Not Stated	Stated	Not Stated
C-1-Time		12		12		12
C-2		12		12	3	9
C-1 Measurement		11		11	1	10
D-1-Numeration		12		12	1	11
D-2		12		12		12
D-3		12		12		12
D-4		12		12		12
D-5		12		12		12
D-1-Place Value		12	4	8		12
D-2		12		12		12
D-2-Addition		12		12		12
D-3		12		12		12
D-4		12		12		12
D-5		12		12		12
D-6		12		12		12
D-7		12		12		12
D-8		12		12		12
D-1-Subtraction		12		12		12
D-2		12		12		12
E-1-Money		12		12		12
E-2		12		12		12
E-3		12		12		12
E-4		12		12		12
E-1-Time		12		12	1	11
E-2		12		12		12
E-3		12		12	1	11
E-4		12		12		12
E-5		12	1	11	1	11
E-1-Measurement		12		12		12
E-2		12	1	11		12
E-3		12		12	1	11
E-4		12		12		12
E-5		12	1	11		12
E-6		12		12	1	11
E-7		12		12	1	11
E-1-Geometry		12		12	1	11
E-2		12		12	2	10
E-3		12	4	8		12
E-4		12		12	1	11
E-5		12		12	1	11
E-6		12		12	1	11

In addition to the data regarding the observability of the action verb and the critical component analysis for each of the objectives the judgement for the critical components for each of the subjects was tallied and compared with the group ratings. That data is included in Table 4.

TABLE 4

NUMBER OF DEVIATIONS FROM
THE GROUP'S RATING ON SPECIFICITY, SIGNAL, AND CRITERION.

	S/G	Signal	Criterion
Cliff Brown	-	-	-
*Tom Comer	-	1	4
Marjorie Devine	-	-	-
*Carol Freidhoffer	-	1	-
*Bill Gibbs	-	3	-
John Jackson	-	-	4
Barbara Kelly	-	-	-
*Roberta Lewis-Lewis	-	1	-
**Grace Owen	-	3	1
*Ardis Rassmus	-	3	9
Nancy Smith	-	-	-
*Edward Stephens	-	1	-

* Several ratings although they deviated from the group rating were judged by experts (Deno and Jenkins) to be acceptable. The * denotes the number of times a subject deviated from the group (yet was judged to be acceptable by an expert).

Inspection of Table 4 enables a determination of the degree to which any subject's ratings were consistent with the group's ratings. Perhaps the best thing that one can say regarding the consistency or reliability of the various raters is that only 1 of the 12 deviates markedly from the group. It is well to note that a deviation does not necessarily indicate incorrectness on the part of the rater. To determine whether or not a particular deviation was reasonable or acceptable, the investigators independently analyzed each of the objectives on which any one rater had deviated. As the table indicates the independent analyses of the judges confirm the possibility of particular

deviations made by individual subjects. That is not to say that a deviation was judged as correct or incorrect, rather that there was enough ambiguity in the form of the objective to allow for alternative categorization or classification.

DISCUSSION

The results indicate that many widely used and recommended behavioral terms do not describe behavior which is as clearly observable as some have suggested. For example, the word "solve," which is frequently used to describe student behavior on various types of problems in mathematics, is only a little more observable than the terms "to be curious," "to think critically," or for that matter "to appreciate." Words such as "identify," "average," "respond to," "group," "regroup," "measure," (from the IPI mathematics continuum) are all at some intermediate position with respect to observability. One is left with the inescapable conclusion that it is inappropriate to suggest that these terms are descriptive of overt human behavior. Rather, they might better be thought of as more invariant descriptions of behavior than the terms "know" and "appreciate," and less invariant descriptions than the terms "write" and "underline."

Apparently words like "identify," "demonstrate," and "apply" are not rated as directly observable because (and this is a logical conclusion) there are a variety of ways in which one might identify or demonstrate or apply. For example, one might identify (which has a mean rating for observability of 3.83) by "pointing to" (1.08) by "saying (1.25), 'that is one'," or by "circling" (1.25).

There is, in fact, some evidence to indicate that the use of words like identify only obscures the observableness of the behavior. The action word "write," for example, was rated as quite observable (1.33), and the word

"identify" has a rather high mean rating (3.83). To state in the objective that the student is "to identify in writing" is to actually produce an intermediate rating for observability (2.17). It is tempting to conclude that as any two action words are combined into a verb phrase observability is reduced by whatever degree the less observable term is rated as unobservable.

Perhaps, as was suggested earlier, curriculum developers with a behavioristic bent have become sensitive to the criticism of trivialness in the objectives which they have written. Subsequently, they may have fallen prey to using only "moderately observable" terms rather than seem unsophisticated in the subject matter, or in the use of the language. Rather than write objectives containing only the verbs "to underline," "to cross out," "to write," "to say" they have opted for what might be called process words such as "identify," "solve," "recognize," "distinguish," "apply," "determine," and "average." At this point the question may be posed to the users of these terms: "What will someone be doing when he is identifying, recognizing, solving, or deducing?" Some curriculum developers (see for example Wallbesser and Gagné) have addressed themselves to this problem. Their solution has been to define "identify" and "distinguish" (not in themselves adjudged observable) in terms of verbs or action words which are given a high rating on observability (in our case mean ratings close to 1.0). Such an approach would likely be a successful one as long as those definitions are always included with the set of objectives. Possibly a more parsimonious approach would be simply to use the action terms in those definitions such as "write," "circle," "underline." However, the former technique (i.e., defining in observable terms) has an advantage in that it suggests a more general or inclusive capability which the learner has acquired, and, consequently, avoids the criticism of trivialness.

It is worth noting that the arithmetic processes so frequently used as action words in IPI mathematics continuum received very consistent ratings on observability (to add 3.08, to divide 3.08, multiply 3.17, to subtract 2.92). The consistency of this rating is particularly remarkable in that each of these process words was presented on a different page, and, most likely, was not an artifact of the sequence or the procedures employed. Consistency among such processes which seem logically equivalent adds credibility to the scale of observability used for this analysis.

Generally, the action words extracted from the IPI objectives are rated in the intermediate ranges of observability. Although the range of IPI verbs encompasses most of the potential range (from 1.3 to 4.3) the median rating given (2.8) indicates that almost as many verbs are rated toward the unobservable pole as toward the observable pole. The rankings given the verbs extracted from the IPI objectives approximate a normal distribution with the majority of the verbs falling in an intermediate range (2.0 - 3.9). (See Table 5.)

TABLE 5

FREQUENCY DISTRIBUTION OF RATINGS
GIVEN IPI VERBS

1.0 - 1.9	2.0 - 2.9	3.0 - 3.9	4.0 - 4.9
12	19	19	7

Twenty-six of the 57 verbs from the IPI objectives were given mean ratings of 3.0 or greater. The action verbs "to solve," "to apply," "to test," "to determine," "to create," "to recognize," and "to deduce," all from the IPI objectives, were given mean observability ratings greater than 4.0.

A note might be made here of the utility of Appendix D (the alphabetical listing of the action words and their rated means of observability). If, as

we have assumed, the statement about the behaviorality of any particular objective is largely determined by the action word for that objective, then the behavioralness of any particular IPI objective can be determined by extracting the action word from that objective and finding its mean rating for observability in the table. Further, if one wished to write objectives, rather than analyze them in terms of observability, it would be possible to use the table as a basis for selecting words which might be expected to yield, invariant usage in the natural language.

Component Classification

An inspection of the classification data leads to the conclusion that nearly every objective sampled from the IPI Mathematics Continuum may be described as "a general behavioral objective with neither a signal nor criterion explicitly stated." The judges unanimously classified only 2 out of the 64 IPI objectives sampled as having a signal or context stated. None of the IPI objectives rated by the subjects contained an explicit criterion of acceptable performance as defined in the set of instructions. Similarly, no objective was classified as specific, by the investigator's criterion. The latter conclusion is not surprising since the pre-requisite condition for an objective to be classified as specific is that the signal or context must be explicitly stated or contained within the objective. As previously mentioned, only two of the objectives contained a signal, and in both cases the signal was a general stimulus situation rather than a specific problem to be responded to.

It is important to note that the results of this analysis with respect to the IPI objectives in no way should be construed as a criticism of the objectives. The task set by the authors was to establish an instrument for analysis within the context of available writings on the construction of

behavioral objectives. That the IPI objectives are not consistent with the form of instructional objectives which was induced from the writings does not necessarily make them unclear or poorly written objectives.

A final statement should be made regarding the number of IPI objectives which were analyzed. For two reasons the investigators did not believe it to be necessary or possible to analyze all of the objectives from the IPI Mathematics Continuum. First, the consistency of the ratings and analysis given by the subjects indicated,as described earlier, that almost all of the IPI objectives were of a particular type with respect to specificity of context and the criterion. Second, the task of classifying the components of a behavioral objective becomes extremely tedious after about 35 objectives. Since little information is provided by additional classifications the decision was made that it would be unnecessary to require the subjects to analyze more than the number of objectives given.

REFERENCES

Commission on Science Education. Science: a process approach. Part 4. Washington, D. C.: American Assn. for the Advancement of Science, 1963.

Gagné, R. M. Analysis of instructional objectives. Teaching machines and programmed learning, II: Data and directions. (Edited by R. Glaser) Washington: National Educational Assn., 1965.

Mager, R. F. Preparing objectives for programmed instruction. San Francisco: Fearon Publishers, 1962.

Mandler, G. and Kessen. The language of Psychology. New York: John Wiley and Sons, Inc., 1959.

Miller, R. B. The newer roles of the industrial psychologist. Industrial Psychology. (Edited by B. Von Gilmer.) New York: McGraw Hill Book Co., 1961. pp. 353-80.

Taber, J. I.; Glaser, R. and Schaefer, H. H. Learning and programmed instruction. Reading, Mass.: Addison-Wesley, 1965.

Wallbesser, H. H. Curriculum evaluation by means of behavioral objectives. Journal of Research in Science Teaching. 1, pp. 296-301, 1963.

APPENDIX A

SAMPLE TEST BOOKLET

This is a study designed to determine the extent to which various words are labels for behavior which is directly observable. For example, most would agree that the verb "to hit" labels behavior which you can see, while the verb "to believe" labels an internal state which cannot be directly observed.

On the subsequent pages you will find a list of action words, or verbs. Your task is to rate each word on a scale from 1 - 5 from most observable to least observable as follows:

Most Observable				Least Observable
1	2	3	4	5
to hit				to believe
to bite				to sympathize

Words such as "to hit" and "to bite" are to be given a rating of 1, while words such as "to believe" and "to sympathize" are rated as 5.

Many of the words may not in your judgement be rated 1 or 5, and these you are to rate as 2, 3, or 4 as you see fit.

Remember, the rating you give is determined by the extent to which you judge it possible to observe the behavior.

PLEASE PLACE THE NUMBER WHICH YOU GIVE AS A RATING IN FRONT OF THE WORD

to recognize

to identify in writing

to take away

to finish

to average

to put on

to repeat orally

to solve

to see

to measure

to find

to analyze

to mark

to discriminate

to read orally

to place

to make

to infer

to write

to feel

to complete

to distinguish

to subtract

to state

to play

to give

to be aware

to supply

to wonder

to create

to understand

to concentrate

to name

to use

to number

to read

to acknowledge

to shade

to discover

to underline

to add

to test

to realize fully

to locate

to select

to utilize

to utilize

to think

to convert

to appreciate

to regroup

to borrow

to draw

to point to

to construct

to count orally

to walk

to be curious

to respond to

to label

to connect

to learn

to choose

to determine

to inquire

to like

to partition

to apply

to remove

to become competent

to think critically

to cover with a card

to say

to develop

to arrange

to draw

to circle

to divide

to perceive

to select

to fill in

to total

to group

to lever press

to match

to identify

to tell what

to reject

to generate

to deduce

to conclude

to summarize

to know

to demonstrate

to check

to order

to round off

to cross out

to perform

to change

to multiply

to line-draw

APPENDIX B

INSTRUCTIONS FOR COMPONENT CLASSIFICATION

Several well known educators (Miller, Gagné, Mager) have agreed that well-stated behavioral objectives should include at least the following characteristics:

1. A specification of the kind of behavior which the learner is expected to do, that is the action.

2. A description of the important conditions under which the behavior will be expected to occur, that is the signal or context for the action to take place.

3. Description of a criterion of how well the learner must perform to have his behavior considered acceptable, that is, some statements of a criterion of acceptability.

For the moment, we will concern ourselves with only the 2nd and 3rd characteristics, the signal and the criterion for acceptability.

First, let us determine if a given objective includes the statement of context (A) or if the writer of the objective has neglected this aspect (B).

Below are examples of two objectives with

CONTEXTS STATED

A. 1. Given the command "say the numbers from 1 to 10", counts orally from one to ten.

2. Presented with pairs of two numbers connected by an addition sign (+), writes the number that is the total of the pair.

(signal underlined)

Below are examples of two objectives with

CONTEXTS NOT STATED

B. 1. Demonstrates mastery of addition facts through sums to 20.

2. Solves addition and subtraction problems.

A second related classification of behavioral objectives involves an objective's degree of specificity. An objective may be either general or specific depending on whether the test condition is explicitly included in the statement. Some examples will clarify this.

SPECIFIC

(test included)

1. Presented with the statement "regroup the following numbers

(test included) $65 = 60 + \_\_\_\_\_ = 50 + \_\_\_\_\_$.

$168 = \_\_\_\_\_ + 68 = \_\_\_\_\_ + 78$",

writes the numbers in the blanks that complete each operation.

2. When given the following problems

(test included)

9 - 4 = _____

90 - 40 = _____

900 - 400 = _____

9000 - 4000 = _____

writes the numbers indicating the difference.

GENERAL

(test not included)

D. 1. When given several states of the union - and told to write their capitol cities, writes their capitols.

2. When presented with a series of names, writes the names in alphabetical order.

An examination of examples A (1., 2.) and B (1., 2.) in terms of the specific-general classification may be helpful.

Look back to each of those examples now and try to determine if you would classify it as general or specific.

According to the definition only example A - 1. is specific since none of the other three explicitly state the test.

By now, you may have discovered a rule of thumb to aid your classification. The rule states "a signal must be stated explicitly if the objective can be classified as specific. If no signal or context is stated the objective is necessarily general. This cannot be interpreted to mean that if a signal is present the objective is specific, only that for an objective to be specific a signal must be present.

The final classification concerns the presence or absence of an explicitly stated criterion. For the purposes of our classification we will consider two kinds of criteria: (a) answer stated and (b) percentage of correct responses. The appropriate criterion statement is contingent on an objective's specificity or generality, that is, the criterion statement for a specific objective is the correct answer, while the criterion statement for a general objective may be either the correct answer or a percentage of correct responses.

E. 1. When given the problem "5 ÷ 2 = _____", writes 7 in the blank.

2. When given the problem "5 ÷ 2 = _____", writes the correct answer in the blank.

3. When given the problem "5 + 2 = _____", writes the correct answer in the blank nine out of ten times.

Both objectives E-1, E-2, and E-3 are specific objectives since all contain the test conditions. However, only E-1 contains an explicit statement of the criterion; that is, the answer 7.

Consider three more onjectives.

F. 1. Counts orally from one to ten.

2. Counts orally from one to ten by saying "one, two, three, four, five, six, seven, eight, nine, ten."

3. Counts orally from one to ten at least six out of seven times.

Each of the above three objectives are general objectives in that the test is not explicitly stated. This is especially easy to determine since none of the three contain a signal which is a necessary requisite for a specific instructional objective.

Remember now that we said in a general objective the criterion may be stated as either the correct answer or as a percentage of correct responses. Which of the above three objectives fit this rule?

Obviously, F-1 does not fall within the category since neither the correct answer is stated, nor is there a statement of percentage of correct responses. Consequently, F-1 falls into the category of criterion not stated.

However, F-2 fits the rule because the correct answer is explicitly stated. Objective F-3 also fits the rule because a statement of percentage of correct responses is contained. Consequently both F-2 and F-3 fall into the category of criterion stated.

Consider two more examples.

G. 1. When presented with five addition problems, writes the correct answer to four out of the five problems.

2. When presented with five addition problems writes the correct answers to them.

Both objectives are general since neither contain the test. However, G-1 contains a criterion statement while G-2 does not.

Now with the sheet marked CATEGORIES FOR OBJECTIVES you should be able to classify objectives on the basis of the following categories:

1. Specific - General

2. Signal: Stated - Not Stated

3. Criterion: Stated - Not Stated

For practice, rate the following objectives.

Obj. 1 Given a set of pictures name the animals contained in the picture.

Now check your rating with the correct answer: You should have checked the following columns: general, signal-stated, criterion-not stated.

Obj. 2 When presented with the following multiplication problems, 5 x 2 = 10, 6 x 12 = 72, and 2 x 100 = 200, writes the underlined answers.

You should have checked the following columns: specific, signal-stated, and criterion-stated.

If you have further questions, raise your hand for help. Otherwise begin rating the objectives on the sheets handed to you.

APPENDIX C

SAMPLE CLASSIFICATION SHEET

Categories for Objectives

jective umber	Specific	General	Signal		Criterion	
			Stated	Not Stated	Stated	Not Stated

APPENDIX D

THE 99 VERBS WITH MEANS AND VARIANCES

PHENOMENA	$\bar{X}$	S^2
to acknowledge	3.58333	1.17424
*to add	3.08333	1.35606
to analyze	4.83333	0.15151
*to apply	4.25000	1.11363
to appreciate	4.91666	0.08333
*to arrange	2.58333	0.62878
*to average	3.41666	0.81060
to be aware	4.91666	0.08333
to become competent	4.75000	0.38636
to be curious	4.83333	0.I5151
*to borrow	3.58333	0.44696
*to change	2.91666	0.99242
*to check	2.25000	1.29545
to choose	2.83333	0.51515
*to circle	1.25000	0.38636
*to complete	3.25000	0.93181
to concentrate	4.83333	0.33333
to conclude	4.83333	0.15151
*to connect	2.66666	1.15151
*to construct	2.25000	1.29545
*to convert	3.91666	1.35606
*to count orally	1.25000	0.75000
to cover with a card	1.00000	0.00000
*to create	4.33333	1.15151
to crossout	1.16666	0.33333
*to deduce	4.83333	0.15151
*to demonstrate	3.16666	0.87878
*to determine	4.33333	0.60606
develop	4.33333	0.42424

PHENOMENA	$\bar{X}$	S^2
to discover	4.75000	0.20454
to discriminate	4.50000	0.63636
to distinguish	4.16666	0.37878
*to divide	3.08333	0.81060
*to draw	2.08333	0.99242
*to draw	2.41666	1.17424
to feel	4.83333	0.15151
*to fill in	1.66666	0.96969
*to find	3.66666	1.69696
to finish	2.58333	0.62878
to generate	4.33333	0.78787
*to give	2.66666	1.33333
*to group	3.25000	0.56818
*to identify	3.83333	0.87378
*to identify in writing	2.16666	1.42424
to infer	4.75000	0.38636
to inquire	3.50000	0.81818
to know	4.91666	0.08333
*to label	1.75000	1.11363
to learn	4.75000	0.38636
to lever press	1.00000	0.00000
to like	4.75000	0.38636
to line-draw	1.00000	0.00000
*to locate	2.66666	0.60606
*to make	2.41666	0.99242
to mark	1.08333	0.08333
*to match	2.33333	0.78787
*to measure	3.08333	1.17424
*to multiply	3.16666	1.24242

PHENOMENA	$\bar{X}$	s^2
*to name	1.58333	0.81060
to number	1.58333	0.44696
*to order	3.08333	0.99242
*to partition	2.91666	0.44696
to perceive	4.83333	0.33333
*to perform	3.00000	1.81818
*to place	1.91666	0.99242
*to play	2.58333	1.71969
to point to	1.08333	0.08333
to put on	1.41666	0.44696
*to read	2.58333	0.81060
*to read orally	1.50000	0.45454
to realize fully	5.00000	0.00000
*to recognize	4.58333	0.44696
*to regroup	3.16666	1.06060
* to reject	2.75000	1.11363
*to remove	1.91666	0.62878
to repeat orally	1.25000	0.38636
*to round off	3.16666	1.42424
*to respond to	3.33333	0.60606
*to say	1.25000	0.75000
to see	3.83333	2.33333
*to select	2.75000	1.47727
*to shade	1.50000	0.81818
*to solve	4.25000	0.93181
*to state	1.75000	1.47727
*to subtract	2.91666	1.35606
to summarize	3.41666	1.17424
*to supply	3.08333	1.35606

PHENOMENA	$\bar{X}$	s^2
*to take away	2.33333	1.15151
to tell what	1.91666	1.17424
*to test	4.33333	0.42424
to think	4.83333	0.33333
to think critically	4.33333	0.33333
*to total	3.00000	1.01818
to underline	1.16666	0.33333
to understand	5.00000	0.00000
*to use	2.91666	1.17424
to utilize	3.50000	1.00000
to walk	1.16666	0.33333
to wonder	4.91666	0.08333
*to write	1.33333	0.42424